**Avian Temporal Occupancy and Climatic Suitability to Determine Core and Transient Species Distributions in the United States**

BIOL 465

12/9/2015

Sara Snell

**Introduction**

The study of species distributions is of great interest in macroecology. One of the principle determinates of species distribution is climatic suitability (Belmaker, 2009). Local diversity can be determined by biotic and abiotic factors, such as the regional species pool, dispersal ability of species, and species-species interactions (Belmaker, 2009). Environmental factors such as mean temperature, precipitation, elevation, and tree cover (NDVI or EVI) could be significant in determining whether or not a species can survive in a habitat and what its optimal climate would be (Coyle, Hurlbert, & White, 2013). Previous studies have found that species with distinct occupancy patterns respond to environmental predictors as determinants of local richness, but it is unclear exactly how (Coyle et al., 2013).

Generally, species can be classified as common or rare in a given area based on their temporal occupancy (Magurran and Henderson 2003). Common species and rare species can be classified into two designations: core and transient (Belmaker, 2009). Core species are characterized by persisting in an area over time, while transient species can be found in an area only occasionally (Coyle et al., 2013). Transient species are more rare in the community and their local abundance is usually a result of the total regional species pool (Supp, 2014). Core species are generally more common, make up more of the community than transient species, are better adapted to their local environment, and are stronger competitors over transient species (Supp, 2014). Core and transient species exhibit different relative-abundance distributions and can be best explained by the occurrence of transient species in the ecosystem (Ulrich & Ollik, 2004). The relative abundance distributions of core and transient species are characterized by their bimodal nature, with transient species on the more rare end and core species on the more common end (Coyle et al., 2013). This phenomenon has been studied in a variety of organisms, from fish to insects to rodents to birds, meaning that the distinct bidmodal species abundance pattern persists across taxa (Belmaker, 2009; Coyle et al., 2013; Supp, 2014; Ulrich & Ollik, 2004). There has been little research done on core and transient species and why they persist in these patterns over space and time, so the relationship between climatic suitability and avian temporal occupancy is of great interest.

A prevailing predictor of species distributions is the abundant center hypothesis (ACH), which states that species will be most abundant in the center of their range and will decline in density as the fringe of the habitat is reached (Manthey et al., 2015). Manthey *et. al* developed this idea to include environmental determinants to determine the optimal range of a species, termed ACHE, and found that species require certain environmental conditions to persist in an area and are more likely to be classified as transient species in sub-optimal areas (Manthey et al., 2015). By combining the knowledge that species have specific environmental determinants and are more abundant in areas at the center of their optimal conditions, it could be possible to predict species occupancy with environmental variables.

Based on previous studies, core species should be in the center of their geographic range where climatic variables are most suitable and species would shift from core to transient as climatic conditions change (Coyle et al., 2013; Manthey et al., 2015). Additionally, species would likely be bounded by a latitudinal gradient if they have a large core habitat because different climatic variables would be correlated with species range depending on their distribution (Hurlbert & Stegen, 2014). The core-transient distribution results obtained should be similar to the distribution results from the species-occupancy matrix (Coyle et al., 2013).

**Methods**

The North American Breeding Bird Survey was the primary data source for species diversity and occupancy data in this study. The surveys began in 1966 and monitor breeding birds across the continent (mbr-pwrc.usgs.gov). They are conducted along roadsides for 400 km routes by a three-minute point count every 800 meters (mbr-pwrc.usgs.gov). Surveyors record all bird species seen along the route, so hypothetically, all bird species that occur in North America would be included. Species name, AOU code, and occupancy were acquired from this data set. These data are comprehensive and provide valuable temporal, spatial, and richness information about land birds in North America.

Climate data came from the WorldClim website, which provides current climate data from about 1950 to 2000 (worldclim.org/current). Mean annual temperature, mean annual precipitation, and national elevation were downloaded at 10m resolution in raster format for North America and cropped to each BBS route (worldclim.org/current). WorldClim assembled its data primarily from the Global Historical Climatology Network and filled in gaps in the data with a variety of other databases (worldclim.org/methods). Elevation data was assembled using the Shuttle Radar Topography Mission database at 1 km resolution (worldclim.org/methods). EVI was acquired from the Coyle *et. al* paper, but future studies would ideally acquire data from the MODIS package in R (Coyle et al., 2013). The current EVI data was 250 m, 16-day resolution, averaged over the summer months of June, July, and August (Coyle et al., 2013). The area of the EVI was attained by selecting a 40 km radius around the starting coordinate of each available BBS route, which limited the total number of routes that could be compared in this study from around 5,000 to 500 (Coyle et al., 2013). These species should still provide a useful baseline of data that can be added to in the future.

First, all of the environmental variables had to be compiled into one dataset and associated with occupancy data for each species and route. This was accomplished using the packages “raster” and “tidyr” in R (Hijmans 2015, Wickham 2015). All code is attached as an appendix. Next, a z-score was calculated for each environmental variable to assess deviation of each species point from the mean. This was accomplished by subtracting the mean at each route from the observed mean for each species, then by dividing the standard deviation of the species mean. Euclidean distance was calculated for each environmental variable to determine the distance between the calculated environmental optimum and the observed mean environmental deviation at each route. Each Euclidean distance was then aggregated by species and divided by the species-specific standard deviation.

For analysis, a linear model was conducted to compare the relationship of univariate and multivariate relationships of environmental variables and species occupancy. Only species with more than four observations were included to minimize error in the analysis. The slope, p value, and R squared value were calculated for each species and the output was stored in a matrix. The univariate models were plotted on an x-y axis and histograms were created to compare all linear models. Additionally, each combination of R squared value and slope for the environmental variables (temperature vs. precipitation, temperature vs. elevation, temperature vs. EVI, elevation vs. precipitation, elevation vs. EVI, and EVI vs. precipitation) were plotted compared to a 1:1 linear relationship to determine how well each species fit the model. Finally, paired t-tests were calculated for each relationship to determine which variable explained more of the observed variation.

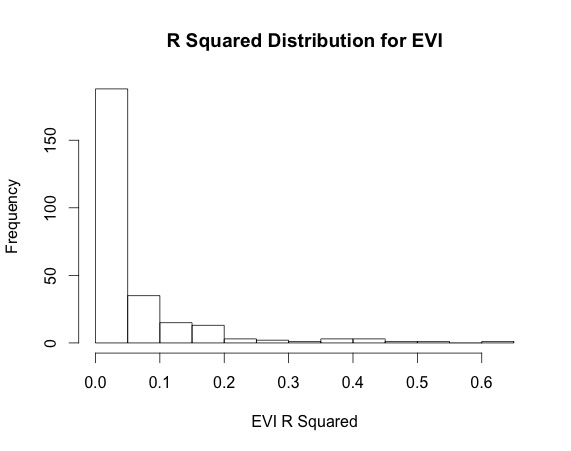
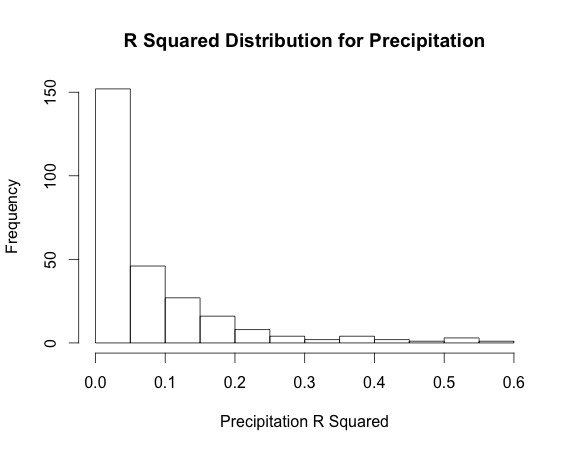
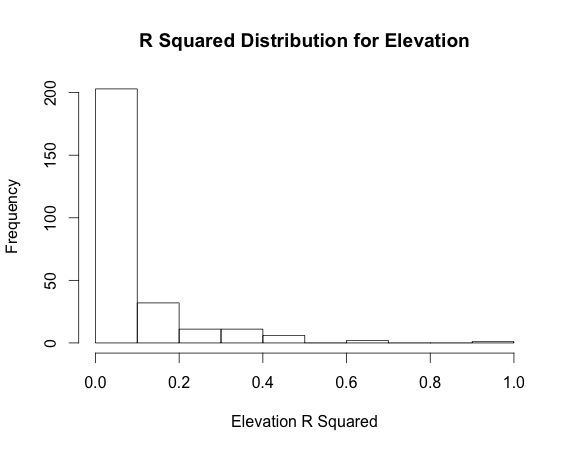
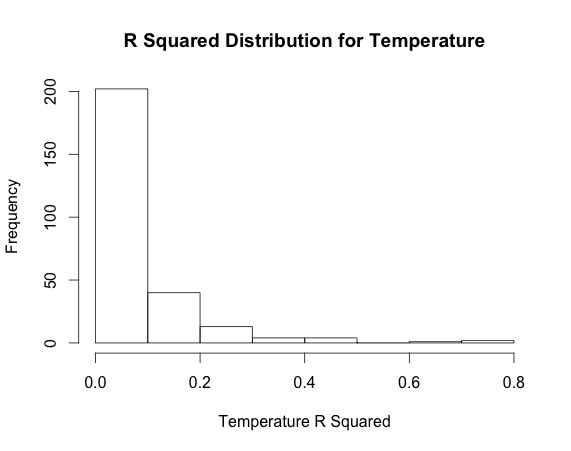
**Results**

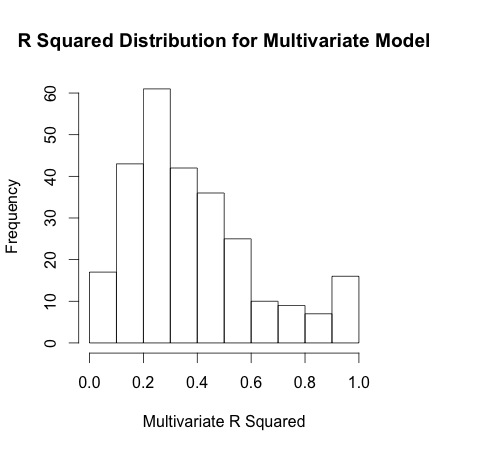
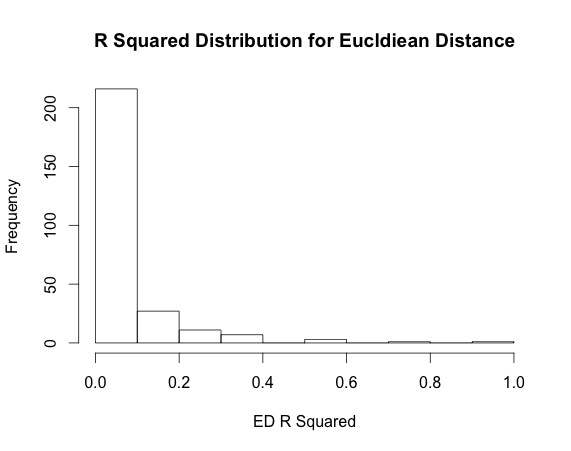
For each of the linear models, histograms were produced to visualize the R squared and slope values (Figure 1). All environmental variable histograms were left-skewed, with temperature deviation from the optimal conditions seeming to have the highest R squared values overall. This indicates that there could be a stronger relationship between temperature deviation and occupancy than the other environmental variables. For Euclidean distance, the majority of R squared values were low, indicating that the model does not fit the data as well as the single environmental variable linear models. This means that the species-specific difference between the observed and expected environmental variables vary widely between species. Finally, the R squared distribution for the multivariate model had the largest variation, indicating that species distributions are likely determined by a unique combination and interaction of environmental variables.

The slope distributions for all variables were centered around zero (Figure 2). The most different environmental variable was EVI deviation, meaning that out of four environmental variables, EVI deviation had a more variable effect on the data, and the strength of the effect varied by species. Ecologically, this indicated that some species likely depend heavily on EVI for assessing climatic suitability whereas for others, it is not a factor. Positive slopes indicated that the models did not accurately predict the species’ optimal climatic conditions, while negative slopes indicated that high species occupancy occurred in areas of high predicted climatic suitability.

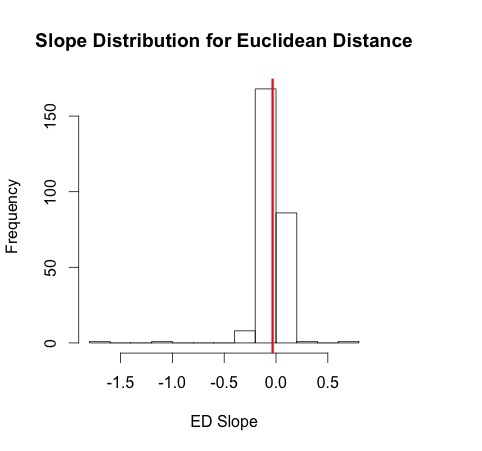
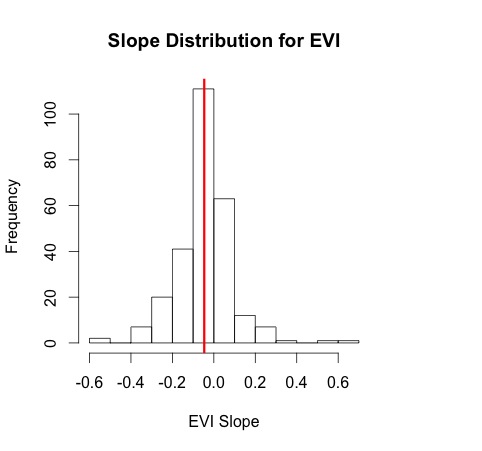
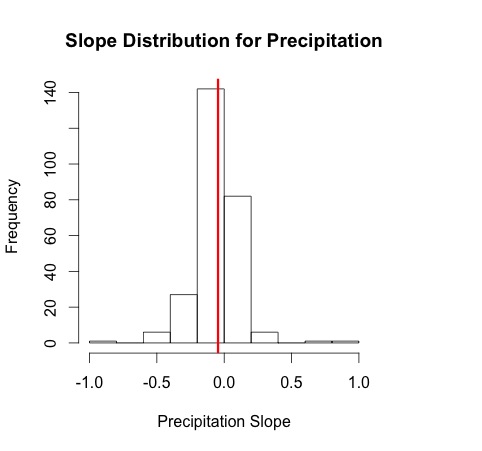
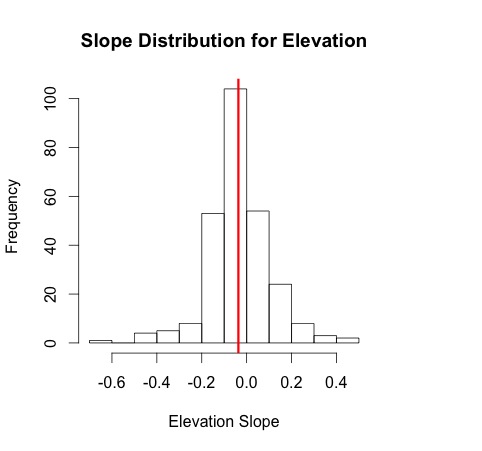
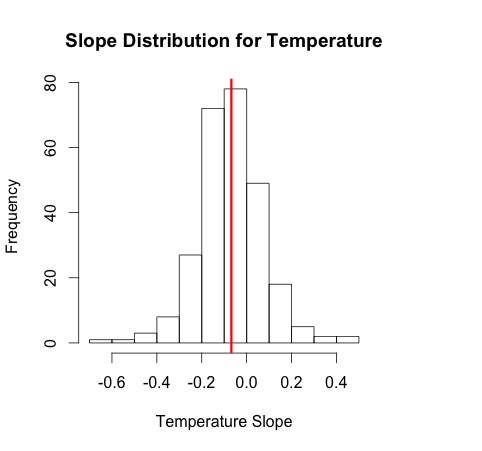
Plots of each univariate linear model were calculated to visualize the relationship of each environmental variable with occupancy by species (Figures 3 through 6). Refer to the appendix for a pdf all plots. Species with negative slopes had more species in the climatic optimal center of their range, with higher occupancies at low environmental variable z-scores. High z-score values indicated a high difference in observed versus optimal climatic conditions, indicating a higher likelihood that the individual observed was near the fringe of their range. Two species were selected to highlight linear models, which were particularly strong at predicting occupancy (Figure 3) and particularly weak (Figure 4), likely a result of these species’ particular life history strategies. Species with the highest R squared values for each climatic variable were plotted, excluding strong relationships with a positive slope (Figure 5). The R squared values ranged from 0.48 to 0.71, indicating that some climatic variables were better fits than others. Additionally, species with the steepest negative slopes for each climatic variable were plotted (Figure 6). These results were summarized in Table 1.

Then, each combination of slope and R squared for environmental variables was plotted. The mean estimate values can be found in Table 2. We can see that temperature deviation led to the steepest negative slope, then precipitation and elevation had about the same value of average estimate, while EVI had the lowest average estimate. When comparing effect sizes, only the temperature deviation R squared points were of interest, with effect size ranging from 2 to 4%. Temperature deviation and EVI deviation had the largest effect size (4.4%), then temperature deviation and elevation deviation (2.9%), and finally temperature deviation and precipitation deviation (2.6%). For both R squared and slope plots, R squared values of 1 were removed because they were the result of lack of data. For slope, similar results were found. For changes in temperature deviation from the optimal conditions, there was a larger response in elevation, EVI, and precipitation when the relationships are compared separately.

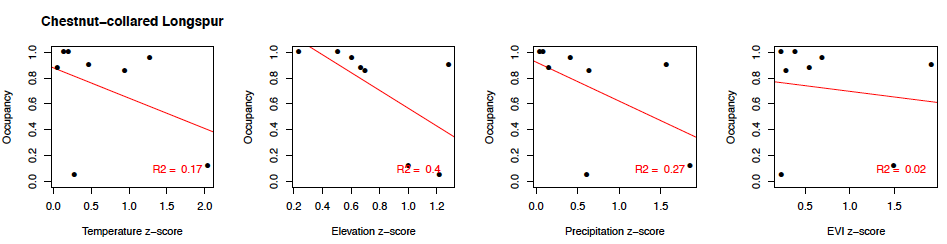
****



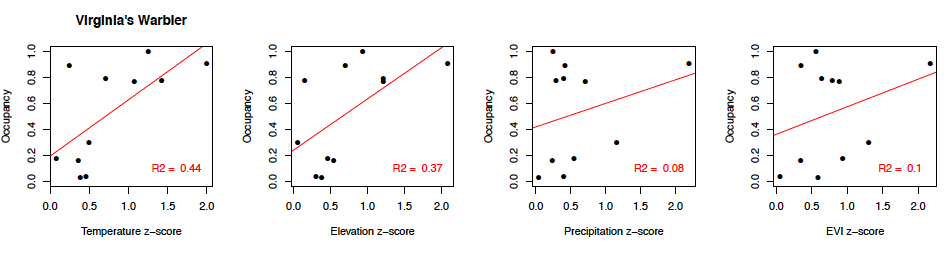
**Figure 1.** R squared distribution histograms for each linear model. The majority of the distributions have low R squared values, with the exception of the multivariate model.



**Figure 2.** Slope distribution histograms for each linear model. Red line is the mean slope for each variable; all variables have means near zero.

****

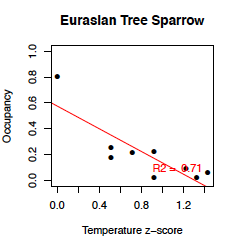
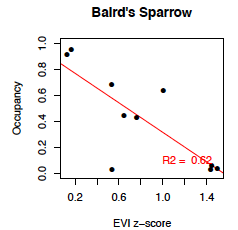
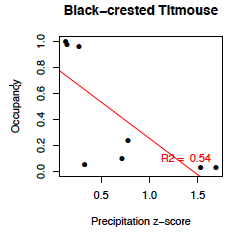
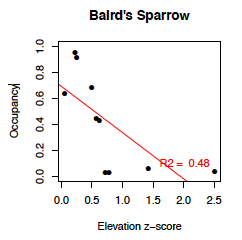
**Figure 3.** Plots of linear model for Chestnut-collard Longspur, a representation of the linear model fitting well for all four environmental variables.

****

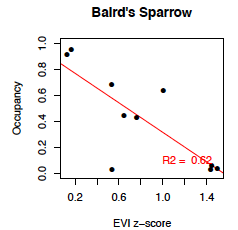
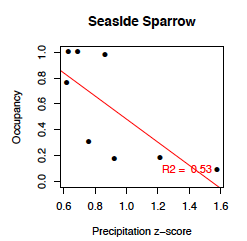
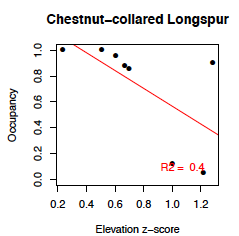
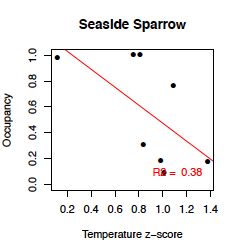
**Figure 4.** Plots of linear model for Virginia’s warbler, which the linear model did not fit well for any environmental predictor variable.

**Table 1.** Species Best Fit by Each Model (only considered models with negative slopes).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Environmental Variable** | **Highest R Squared** | **R sq.**  **Value** | **Best Fit Slope** | **Slope Value** |
| Temperature | Eurasian Tree Sparrow | 0.71 | Seaside Sparrow | -0.69 |
| Elevation | Baird’s Sparrow | 0.48 | Chestnut Collared Longspur | -0.36 |
| Precipitation | Black-crested Titmouse | 0.54 | Seaside Sparrow | -0.89 |
| EVI | Baird’s Sparrow | 0.62 | Baird’s Sparrow | -0.57 |

**Figure 5.** Highest R squared values for Temperature, Elevation, Precipitation, and EVI by species.



**Figure 6**. Steepest slopes for Temperature, Elevation, Precipitation, and EVI by species.

**Table 2.** Ranked mean estimates of each environmental variable. A steeper negative slope indicates higher species occupancy at predicted optimal conditions.

|  |  |
| --- | --- |
| **Variable** | **Mean Estimate** |
| Temperature | -0.0680 |
| EVI | -0.0463 |
| Precipitation | -0.0457 |
| Elevation | -0.0368 |
| Euclidean Distance | -0.0330 |

**Discussion**

It was predicted that core species would occur in the center of their geographic range where climatic variables were ideal and species would shift from core to transient as conditions became less optimal. Additionally, core species should be more successfully predicted by local environmental variables than transient species (Coyle et al., 2013). Overall, this seemed to be the trend but the results were more species-dependent than expected. Mid-range species occurring primarily in the continental United States were best predicted, while narrow specialists and common generalists seemed to be constrained by additional factors. Other authors have found there to be complications with rare species because it is difficult to know whether their life histories guarantee rarity in their environment or if they are transient in one area and core in another area (Coyle et al., 2013; Magurran & Henderson, 2003).

Temperature had the greatest predicting potential, supporting the hypothesis that a latitudinal gradient could be evident when studying species. EVI and precipitation had the second highest predictive powers, while elevation had the least. This seemed to indicate that more species are bounded by temperature and some have either separate or additional requirements of tree cover and precipitation, while few species require specific elevations as determinants of occupancy.

Additionally, the linear models were successful predictors of occupancy for species with environmental requirements. The models were unsuccessful at predicting species that occurred primarily outside of the study range as well as the most common and most rare species. For example, the purple finch was not predicted well by the model, likely because its range is large and is likely bounded by non-environmental determinates. It was difficult to tease apart distinct differences between core and transient species because many widespread species that were not fit well by the model could be considered core in the majority of the Unites States and their habitats were not likely determined by environmental factors. The multivariate analysis of Euclidean distance seemed to have the lowest predictive ability compared to the univariate models based the histograms and mean estimates. This made it difficult to determine the precise combination of environmental factors that determine a species range. The additive multivariate linear model R squared value was so variable because each species likely required unique environmental variables as determinants of their optimal environment, which did not seem to be predicted well with Euclidean distance.

One issue with the data was that the BBS only covered the continental US and parts of Canada, which distorted the ranges of species who frequently move between the US and Mexico and/or Canada (mbr-pwrc.usgs.gov). One example observed was Bendire’s Thrasher, which primarily exists in Mexico, but has some core presence in the Southwest. This led to positive slopes for each environmental variable predictor, which would indicate that as occupancy increased, the climatic conditions became less suitable and that the model did not fit the data. To improve the data in future studies, species whose primary range occurred outside the continental United States should be eliminated from the dataset to reduce error in the linear models. The BBS could also be supplemented with valid data from a reputable data source such as eBird or the BBC to determine if a more complete occupancy dataset could be obtained, particularly for rare avian species. Perhaps avid birders would seek out rare birds more successfully than routine BBS surveys, so with other datasets a more complete picture of species occupancy in North America could be developed.

One potential reason for lack of fit in the linear models was that the environmental relationships predicted were not a good fit. This could be due a variety of reasons such as the nature of data mentioned above, species responding to non-environmental determinants, or changing species distributions as a result of human activity (Manthey et al., 2015). Other studies hypothesized that abnormal occupancy-environmental relationships (i.e., species not being most abundant at their environmental abundant center) could be due to local landscape processes such as urbanization, deforestation, and habitat fragmentation (Manthey et al., 2015). Humans could also play a more direct role in disturbing a habitat with noise pollution, hunting, or introducing diseases and competitors (Manthey et al., 2015). All of these scenarios could play a role in the models not fitting the data for a particular species, so more studies would have to be done to understand the particular needs of each species.

Future studies should attempt to better understand core and transient species distributions and how the environment determines those relationships. One way to improve the study would be to incorporate biotic factors such as competition and predation to gain a more comprehensive understanding of species-specific requirements for range and how biotic and abiotic factors relate. Additionally, anthropogenic alterations to the environment would be interesting to include such as quantifying habitat connectivity, urbanization, and potential future scenarios for climate change. As humans continue to alter the environment, it becomes more critical to understand occupancy patterns in avian species and what factors determine those patterns.

**References**

Belmaker, J. (2009). Species richness of resident and transient coral-dwelling fish responds differentially to regional diversity. *Global Ecology and Biogeography*, *18*(4), 426–436. http://doi.org/10.1111/j.1466-8238.2009.00456.x

Coyle, J. R., Hurlbert, & White, E. P. (2013). Opposing Mechanisms Drive Richness Patterns of Core and Transient Bird Species. *The American Naturalist*, *181*(4), E83–E90. http://doi.org/10.1086/669903

Hurlbert, A. H., & Stegen, J. C. (2014). On the processes generating latitudinal richness gradients: identifying diagnostic patterns and predictions. *Evolutionary and Population Genetics*, *5*, 420. http://doi.org/10.3389/fgene.2014.00420

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*(6933), 714–716. http://doi.org/10.1038/nature01547

Manthey, J. D., Campbell, L. P., Saupe, E. E., Sobern, J., Hensz, C. M., Myers, C. E., … Barve, V. (2015). A test of niche centrality as a determinant of population trends and conservation status in threatened and endangered North American birds. *Endangered Species Research*, *26*(3), 201–208. http://doi.org/10.3354/esr00646

Supp, S. (2014). Species-level and community-level responses to disturbance: a cross-community analysis. *Ecology*, *95*, 1717–1723.

Ulrich, W., & Ollik, M. (2004). Frequent and occasional species and the shape of relative-abundance distributions. *Diversity and Distributions*, *10*(4), 263–269. http://doi.org/10.1111/j.1366-9516.2004.00082.x

North American Breeding Bird Survey: http://www.mbr-pwrc.usgs.gov/bbs/genintro.html, accessed 11/6/15.

WorldClim Climate Data: http://www.worldclim.org/current, accessed 11/6/15.

Robert J. Hijmans (2015). raster: Geographic Data Analysis and Modeling. R package version 2.4-20. http://CRAN.R-project.org/package=raster

Hadley Wickham (2015). tidyr: Easily Tidy Data with `spread()` and `gather()` Functions. R package version 0.3.1. http://CRAN.R-project.org/package=tidyr